Online Appendix

A Supplementary Results

A.1 RMSE for Estimator and Model Comparison

In this section we report some additional statistical results. We begin with the RMSE[1] for all the models in Table 1. It is immediately clear that the LDV model performs much worse than all the other methods of estimation under this set of experimental conditions (except OLS with a second lag of Y_t). The RMSE for the LDV model is around 0.28 while for the other models the RMSE was typically 0.10. The RMSE for the LDV model is the same regardless of whether the DGP is static or weakly dynamic. The ARMA model has the lowest RMSE; however, the difference between the ARMA model and the GLS models is trivial. We also see that the effect of the omitted variable bias in the OLS model without an LDV is clearly discernible in the RMSE. While the RMSE for OLS without an LDV is quite low when the DGP is static, the RMSE triples from 0.16 when the model is static to 0.46 once α is 0.50.

Table 1: Comparative Root Mean Squared Error (RMSE) For Static to Weakly Dynamic DGP

	α			
Model	0.00	0.10	0.20	0.50
LDV	0.284	0.298	0.287	0.282
ARMA	0.097	0.093	0.101	0.122
GLS-Cochrane	0.099	0.097	0.103	0.149
OLS[a]	0.163	0.177	0.177	0.456
GLS-Prais	0.098	0.100	0.102	0.148
2LDV	0.283	0.294	0.282	0.287

RMSE for β, Coefficient of X_t.
See Appendix For Model Details.
Results are based on 1000 Monte Carlo
replications.
ϕ: 0.75; ρ: 0.95
[a] Does not include any lags.

Table 2 contains the RMSE for when the DGP is dynamic. We start with the RMSE for the situation in which there is no residual autocorrelation in the error term. Under this condition, OLS with an LDV should be consistent. While we do not assess the fit of the LDV model improves as the sample size increases in this experiment, the LDV model performs well even in this relatively small sample (100). The LDV model has the lowest RMSE of all the models estimated, being *much* lower than that of OLS without an LDV and the two GLS models (0.07 compared to 1.0 and 0.45, respectively). The question, though, is whether the LDV model provides superior estimates once the error term is no longer IID?

---

[1]More specifically, the RMSE is: $\sqrt{\sum_{l=1}^{1000}(\hat{\theta}-\theta)^2}$. The RMSE calculation, here, only includes β the parameter associated with the X_t variable.

Table 2: Comparative RMSE For Dynamic DGP

			ϕ	
Model	0.00	0.10	0.20	0.50
LDV	0.071	0.072	0.076	0.126
ARMA	0.123	0.122	0.122	0.128
GLS-Cochrane	0.447	0.420	0.391	0.296
OLS[a]	1.033	1.035	1.037	1.056
GLS-Prais	0.453	0.427	0.398	0.302
2LDV	0.073	0.073	0.083	0.163

RMSE for β, Coefficient of X_t.
See Appendix For Model Details.
Results are based on 1000 Monte Carlo
replications.
α: 0.75; ρ: 0.95
[a] Does not include any lags.

While the analytic results tell us that OLS is inconsistent, the substantive size of the problem appears to be relatively small. Specifically, the RMSE for the LDV model increases only slightly (from 0.071 to 0.072) once a small amount of autocorrelation is introduced into the error term. It is only when the error term is strongly autoregressive, $\phi = 0.50$, that the ARMA model approximates the performance of OLS with an LDV. The GLS models and OLS without an LDV are clearly dominated by the other models. The RMSE for the two GLS models (Cochrane and Prais) is, on average 5-6 times higher than that of the LDV model. The RMSE for OLS without an LDV is nearly 30 times larger than the LDV model. The omission of dynamics from these models induces serious specification bias.

A.2 Full Rejection Rate Tables

Here we include the full results from the tests for the rejection rates. Table 3 contains the percentage of times that we conclude that X_t has no statistically significant effect on Y_t despite the true effect being 0.50 (the percentage of times we make a Type II error).

A.3 Full Results for Total Effect

In Table 4 we present the bias in the long-run multiplier or total effect of X_t on Y_t under all the conditions we used in the second Monte Carlo analysis. Generally, the bias is larger than that of the initial effect, particularly when ϕ is 0.50.

B Stationarity Proof

We prove that the sum of α and ϕ must be less than the absolute value of 1 for the model to be stationary. As usual we start with the following DGP:

Table 3: Rejection Rate For $\hat{\beta}$, The Coefficient Of X

		$\phi = 0.00$			
N	ρ	0.65	0.75	0.85	0.95
25		13.90	9.80	8.20	05.60
50		0.70	1.40	0.30	0.00
75		0.00	0.10	0.00	0.00
100		0.00	0.00	0.00	0.00
250		0.00	0.00	0.00	0.00
500		0.00	0.00	0.00	0.00
1000		0.00	0.00	0.00	0.00
		$\phi = 0.10$			
N	ρ	0.65	0.75	0.85	0.95
25		16.90	14.20	10.30	9.20
50		1.30	1.50	0.40	0.30
75		0.10	0.00	0.00	0.00
100		0.00	0.00	0.00	0.00
250		0.00	0.00	0.00	0.00
500		0.00	0.00	0.00	0.00
1000		0.00	0.00	0.00	0.00
		$\phi = 0.20$			
N	ρ	0.65	0.75	0.85	0.95
25		17.80	14.80	10.00	11.30
50		1.80	1.30	0.60	0.70
75		0.60	0.00	0.00	0.00
100		0.00	0.00	0.00	0.00
250		0.00	0.00	0.00	0.00
500		0.00	0.00	0.00	0.00
1000		0.00	0.00	0.00	0.00
		$\phi = 0.50$			
N	ρ	0.65	0.75	0.85	0.95
25		28.5	22.00	20.20	20.00
50		8.20	6.30	4.70	2.70
75		2.40	1.70	1.10	0.70
100		0.50	0.50	0.10	0.00
250		0.10	0.00	0.00	0.00
500		0.00	0.00	0.00	0.00
1000		0.00	0.00	0.00	0.00

Results are based on 1000 Monte Carlo replications. Cell entries represent the percentage of time we fail to reject $\hat{\beta} = 0$.

Table 4: Bias, As A Percentage, In The Long Run Multiplier Effect of X_t

		$\phi = 0.00$			
N	ρ	0.65	0.75	0.85	0.95
25		−6.07	−1.41	−6.77	−5.85
50		−5.57	−4.02	−2.20	−2.24
75		−4.55	−3.21	−2.22	−2.33
100		−3.77	−3.09	−1.44	−1.25
250		−1.12	−0.18	−0.85	−0.38
500		−0.05	−0.21	−0.09	−0.09
1000		−0.53	−0.06	0.31	−0.11
		$\phi = 0.10$			
N	ρ	0.65	0.75	0.85	0.95
25		0.96	−0.89	−0.61	−0.38
50		3.99	1.36	1.35	−0.01
75		3.69	2.90	2.01	1.47
100		4.73	1.49	2.07	0.71
250		6.07	4.58	2.33	1.21
500		6.26	4.46	2.67	0.91
1000		6.54	4.45	2.52	0.85
		$\phi = 0.20$			
N	ρ	0.65	0.75	0.85	0.95
25		−28.66	27.66	12.04	4.69
50		12.68	12.33	6.82	3.33
75		12.62	10.91	5.38	3.84
100		13.63	10.32	6.44	2.49
250		14.73	11.04	6.34	2.24
500		14.39	11.13	6.53	2.37
1000		14.98	11.01	6.22	2.07
		$\phi = 0.50$			
N	ρ	0.65	0.75	0.85	0.95
25		15.68	144.78	39.92	91.08
50		11.16	52.89	39.03	26.12
75		67.95	50.41	34.71	18.76
100		66.76	49.20	32.02	16.10
250		64.25	46.80	28.46	10.85
500		63.11	45.32	27.40	9.56
1000		62.61	45.15	26.44	8.31

Results are based on 1000 Monte Carlo replications.
Cell entries represent the average percentage
of bias in the long run effect of X_t.
Long-Run Effect: $\frac{\beta}{1-\alpha}$

$$\begin{aligned}
y_t &= \alpha y_{t-1} + \varepsilon \\
\varepsilon_t &= \phi \varepsilon_{t-1} + v_t
\end{aligned} \tag{1}$$

The addition of an explanatory variable makes no difference to the proof, so we leave it out for clarity. With the above equations, we can substitute in equivalent terms and rearrange to produce the following time series process:

$$\begin{aligned}
y_t - \alpha y_{t-1} &= \phi(y_{t-1} - \alpha y_{t-2}) + v_t \\
y_t &= (\phi + \alpha)y_{t-1} - \phi\alpha y_{t-2} + v_t
\end{aligned} \tag{2}$$

The DGP is clearly an AR 2 process and the sum of $(\phi + \alpha)$ and $-\phi\alpha$ must be less, in absolute value, than 1 for the process to be stationary given the stationarity conditions for an AR process.

C Estimator Notation and Details

Here, we provide a more detailed overview of the estimators we use in the model comparison section of the paper. We, first, outline the Cochrane-Orcutt and Prais-Winsten estimators simultaneously since the two are closely related. The reader should note that Cochrane-Orcutt is asymptotically equivalent to estimating a model with AR(1) errors via Non-Linear Least Squares, and Prais-Winsten is asymptotically equivalent to estimation via full information maximum likelihood. For both procedures with a single covariate, we estimate with OLS:

$$y_t = ax_t + c + u_t \tag{3}$$

An estimate of ρ, the correlation in the residuals, is obtained by estimating the following auxiliary regression:

$$u_t = \rho u_{t-1} + e_t \tag{4}$$

We apply a Cochrane-Orcutt transformation to equation 3, and the estimate of ρ is used to now

estimate:

$$y_t - \rho y_{t-1} = a(x_t - \rho x_{t-1}) + c(1 - \rho) + e_t \tag{5}$$

Clearly in equation 5 observation 1 will be lost. The transformation for $t = 1$ is:

$$\sqrt{1 - \rho^2 y_1} = a(\sqrt{1 - \rho^2 x_1}) + c\sqrt{1 - \rho^2} + \sqrt{1 - \rho^2 e_t} \tag{6}$$

The difference between Cochrane-Orcutt and Prais-Winsten is Cochrane-Orcutt does not transform the first observation as in equation 6, and it is lost. As such, Cochrane-Orcutt regression is somewhat less efficient than Prais-Winsten. The above steps are iterated until the estimate of ρ is within a specified tolerance.

The other estimator we used was an ML routine for $\text{ARMA}(p,q)$ models. Here, the likelihood is computed via a state space representation of the ARMA process; innovations and their variances are found with a Kalman filter. See Hamilton (1994) for more details. The notation for the estimator, while not complex, is fairly lengthy, but is standard in most statistical software packages. Finally, we used OLS with varying numbers of lags of the dependent variable on the right-hand side. In the model comparison section, we compare OLS without any lags of Y_t on the right-hand side to models with one or two lags of Y_t on the right-hand side. Each time, however, OLS is applied to the equation regardless of the number of lags included in the estimating equation.

Finally, why not include quasi-instrumental variables among the estimators? First, we don't want to encourage the use of two-stage least squares. Second, the quasi-instruments approach is particularly problematic given our research design. In choosing an instrument for the lag of Y, we could either use lags of X or some Z variable. To make either one of these variables into quasi-instruments would require making the instrument correlated with the error term of Y. Such a correlation would of course introduce some bias into the estimates. It would then be impossible to know how much of the bias was caused by the correlated errors as opposed to the autocorrelated errors. Moreover, we would have to make arbitrary decisions about the level of correlation between the instrument and the error term. This would introduce another parameter to vary making the

quasi-instrument estimates non-comparable to the other estimates. Moreover, the general reason to use quasi-instruments is for gains in efficiency at the cost of some bias. Here, efficiency is not the question, so we see no reason to include a quasi-instruments approach.

D   The Effect of Model Fit

How well a model fits can also affect the overall performance of an estimator. To investigate how the model fit affected the performance of LDV models, we replicated the Monte Carlo analysis under three additional conditions. Here, we varied the value of β. The first value we used for β was 0.00. The reason we set β to 0.0 for the first condition differs from just assessing the effect of model fit. While the problem of spurious correlations generally does not occur with stationary data (as the data here are), in small sample sizes the problem of a spurious relationship can occur when X_t and Y_t both trend for the sample under observation. A lagged dependent variable can alleviate this problem. By setting β to 0.00, we can observe whether such spurious correlations occur when the residuals are autocorrelated. In Table 5, we report the bias but not as a percentage, since we can't divide by zero. Clearly, with the lag, one fails to find much in the way of spurious effects, since even when X_t is highly autoregressive, the estimated effect is zero.

In the second condition, we set β to 0.10 and in the third to 1.50. Here, we can observe if the size of the effect of X_t on Y_t matters. The results from these analyses are in Tables 6 and 7. In general, the results mirror those reported earlier. If the value of ϕ is 0.00 the estimates will be highly precise as the sample size increases regardless of how strong the effect of X_t is on Y_t. Not surprisingly, strong effects are more impervious to the bias present in LDV models. Smaller effects are always harder to find and the bias exacerbates this to some extent.

The overall fit of the models did not, except by sample size, vary much. Moreover, the overall model fit as summarized by the adjusted R^2 was not related to the amount of bias in the model. For the models where $\beta = 1.5$, the smallest average adjusted R^2 was .90 for 25 cases and was generally higher, even when the bias was fairly high under the $\phi = 0.50$ condition. When β was 0.10 the adjusted R^2 often was quite low even when the bias was quite small. For example, when ϕ was 0.00 and the bias very small, the adjusted R^2 was as small as .38 for 25 cases and grew to .57 for

Table 5: Bias in $\hat{\beta}$, the coefficient of X_t With β set to 0.00.

		$\phi = 0.00$			
N	ρ	0.65	0.75	0.85	0.95
25		0.01	0.001	0.008	-0.003
50		-0.002	0.0009	-0.0002	0.003
75		-0.003	-0.0007	-0.00007	-0.001
100		-0.002	-0.0009	-0.001	-0.002
250		0.003	0.006	-0.001	-0.0008
500		0.002	0.002	0.0008	0.00002
1000		-0.0005	0.00002	-0.0005	-0.00004
		$\phi = 0.10$			
N	ρ	0.65	0.75	0.85	0.95
25		-0.002	-0.007	-0.01	-0.009
50		-0.002	-0.002	0.00008	0.0003
75		-0.0001	-0.002	0.003	0.004
100		-0.0003	-0.008	0.0002	-0.0004
250		0.0005	0.002	-0.0002	0.0007
500		0.0005	0.0008	0.0003	0.0002
1000		-0.000007	-0.0009	-0.0005	0.00005
		$\phi = 0.20$			
N	ρ	0.65	0.75	0.85	0.95
25		0.004	0.0006	0.01	-0.004
50		-0.0009	0.006	0.00002	0.001
75		0.002	0.005	-0.002	0.001
100		-0.003	-0.0008	-0.002	-0.0009
250		0.0009	0.001	0.0005	-0.0006
500		-0.001	0.002	0.001	0.0009
1000		-0.0006	0.001	-0.00005	0.0004
		$\phi = 0.50$			
N	ρ	0.65	0.75	0.85	0.95
25		-0.02	0.005	0.02	-0.005
50		0.005	-0.0007	-0.007	0.004
75		-0.004	-0.003	-0.001	0.0001
100		0.002	-0.002	-0.003	0.0009
250		-0.002	0.002	-0.0009	0.00004
500		-0.0005	-0.0003	-0.00009	0.00008
1000		-0.002	-0.0009	0.0001	-0.0003

Results are based on 1000 Monte Carlo replications. Cell entries represent the average bias in the estimated coefficient.

Table 6: Bias in $\hat{\beta}$, the coefficient of X_1 With β set to 0.10

			$\phi = 0.00$		
N	ρ	0.65	0.75	0.85	0.95
25		23.62	28.24	29.34	26.96
50		5.62	10.83	13.75	25.38
75		2.46	7.02	10.04	15.07
100		2.60	5.36	7.40	9.56
250		4.50	3.80	1.91	4.79
500		2.85	3.29	2.70	2.82
1000		-0.67	0.59	0.24	1.41
			$\phi = 0.10$		
N	ρ	0.65	0.75	0.85	0.95
25		7.79	5.48	7.26	14.27
50		5.76	3.66	7.52	10.75
75		0.81	-1.32	6.21	8.18
100		-0.74	-7.89	0.08	0.19
250		-2.48	-1.58	-5.15	-5.70
500		-3.38	-4.51	-6.48	-9.27
1000		-4.47	-6.79	-8.32	-10.53
			$\phi = 0.20$		
N	ρ	0.65	0.75	0.85	0.95
25		-8.40	-9.46	-10.74	-12.20
50		1.14	0.67	-0.09	0.18
75		4.11	4.23	3.29	2.95
100		5.88	5.79	5.58	4.36
250		8.27	8.14	7.93	7.72
500		9.16	8.99	8.87	8.34
1000		9.59	9.44	9.24	8.91
			$\phi = 0.50$		
N	ρ	0.65	0.75	0.85	0.95
25		-14.02	7.55	21.99	-6.12
50		-2.55	-10.99	-20.27	-15.34
75		-14.89	-17.83	-22.18	-28.50
100		-11.22	-19.99	-26.58	-30.99
250		-19.27	-21.15	-32.48	-42.90
500		-19.09	-25.31	-34.07	-46.76
1000		-21.49	-26.88	-35.03	-49.05

Results are based on 1000 Monte Carlo replications.
Cell entries represent the average percentage of bias in
the estimated coefficient.

Table 7: Bias in $\hat{\beta}$, the coefficient of X_1 With β set to 1.50

N	ρ	0.65	0.75	0.85	0.95
		$\phi = 0.00$			
25		2.12	2.56	2.83	2.24
50		0.79	0.94	0.92	1.39
75		0.43	0.59	0.53	0.94
100		0.33	0.57	0.32	0.31
250		0.38	0.24	0.19	0.15
500		0.21	0.27	0.11	0.10
1000		0.06	0.00	0.09	0.13
		$\phi = 0.10$			
N	ρ	0.65	0.75	0.85	0.95
25		1.39	1.24	1.33	1.20
50		0.55	0.30	0.77	0.83
75		0.21	−0.09	0.48	0.32
100		−0.01	−0.38	0.03	−0.05
250		−0.35	−0.23	−0.37	−0.38
500		−0.49	−0.42	−0.46	−0.43
1000		−0.56	−0.66	−0.59	−0.48
		$\phi = 0.20$			
N	ρ	0.65	0.75	0.85	0.95
25		1.70	1.32	1.85	1.42
50		0.27	0.55	0.23	0.33
75		−0.05	−0.09	−0.53	−0.17
100		−0.87	−0.62	−0.88	−0.41
250		−1.05	−1.05	−1.04	−0.95
500		−1.24	−1.16	−1.15	−1.11
1000		−1.28	−1.22	−1.32	−1.15
		$\phi = 0.50$			
N	ρ	0.65	0.75	0.85	0.95
25		1.69	1.31	1.85	1.42
50		0.26	0.55	0.23	0.33
75		−0.05	−0.09	−0.53	−0.18
100		−0.87	−0.62	−0.88	−0.41
250		−1.05	−1.05	−1.04	−0.95
500		−1.24	−1.16	−1.15	−1.11
1000		−1.28	−1.22	−1.32	−1.16

Results are based on 1000 Monte Carlo replications.
Cell entries represent the average percentage of bias
in the estimated coefficient.

500 cases. So in general, the overall model fit was roughly indicative of how well OLS estimated the coefficients.

References

Hamilton, James D. 1994. *Time Series Analysis.* Princeton, NJ: Princeton University Press.